Methods for evaluating model performance of IMDAA

Forecasting Research
Technical Report No: 594

September 15, 2014

Sana Mahmood, Peter Jermey and Richard Renshaw

Contents

1 Introduction **2**

2 Reanalysis system **2**

3 Validation data **6**

3.1 Independent observations 6

3.2 Reanalyses 6

4 Assessment methods **7**

4.1 Precipitation 7

4.2 Monsoon onset, peak, withdrawal and duration 9

4.3 Monsoon variability and teleconnection links 9

4.4 Reanalysis Quality 12

4.4.1 Points-based Validation 14

4.4.2 Spatial Validation 15

5 Conclusions **15**

A Points-Based Validation **18**

A.1 Categorical 18

B Fractions Skill Score **19**

© Crown Copyright 2014

1 Introduction

The Indian Monsoon Data Assimilation and Analysis (IMDAA) project is a formal collaboration between the Met Office (MO), the National Centre for Medium Range Weather Forecasting (NCMRWF) and the India Meteorological Department (IMD). A principal aim of this four year project is to develop and run a regional reanalysis over the Indian subcontinent. The reanalysis will produce a consistent dataset of high resolution fields for a wide range of atmospheric variables.

The monsoon is the primary weather phenomenon affecting the Indian subcontinent and is distinguished by the seasonal reversal of wind and the associated changes in precipitation. There are several comprehensive reviews of the monsoon that describe its main characteristics, predictability and prediction (Webster et al., 1998; Gadgil, 2003; Goswami, 2004). Since the monsoon provides around 80% of annual rainfall (Turner and Annamalai, 2012) agriculture in the region is highly dependent on the monsoon's strength and onset date. Climate models yield uncertainty about how the changing climate is affecting the monsoon (Dobler and Ahrens, 2011) and therefore the IMDAA reanalysis will be a useful tool for increasing our understanding of the monsoon and how it has changed over the past four decades.

The prediction of the monsoon is notoriously difficult and there are many aspects of monsoon processes from the onset through the development and decay that are relatively poorly understood and represented in model simulations (Turner et al., 2011). The IMDAA reanalysis will produce a long-term historical record of climate and extreme weather events over a region spanning the Indian peninsula and surrounding areas in the form of high-resolution datasets. To be able to use this data confidently in studies it is paramount that the quality of the reanalysis is well understood.

The roles of the participating organisations in this collaboration consist of the Met Office providing scientific expertise to accompany the use of the UM for tailored regional studies of the Indian monsoon. NCMRWF will take the lead role in performing numerical simulations of past monsoons on Indian HPC's, and ultimately ownership of the IMDDAA data assimilation configuration as a vehicle for future monsoon studies by Indian scientists. IMD will be responsible for providing the link to local Indian observation datasets, including local data recovery efforts of archived records.

This report aims to identify the diagnostic tools and measures that could be used to give a thorough assessment of the quality of the IMDAA reanalysis dataset. This is not an exhaustive list however, since exploration of the data after production may lead to further methods of assessment. This report contains preliminary ideas to provoke further discussion and investigation. The report is organised as follows: Section 2 describes the reanalysis system proposed for the IMDAA project, Section 3 outlines the various data available that could be used for validation of the reanalysis. Finally, assessment methods are proposed in Section 4, followed by conclusions in Section 5.

2 Reanalysis system

The reanalysis system will be based on the Met Office four-dimensional variational assimilation scheme, 4DVAR (Rawlins et al., 2007), and its Unified Model, UM (Davies et al., 2005). The system will assimilate all suitable observations including surface observations, radiosondes, satellites and aircraft as detailed in Table 1.

© Crown Copyright 2014

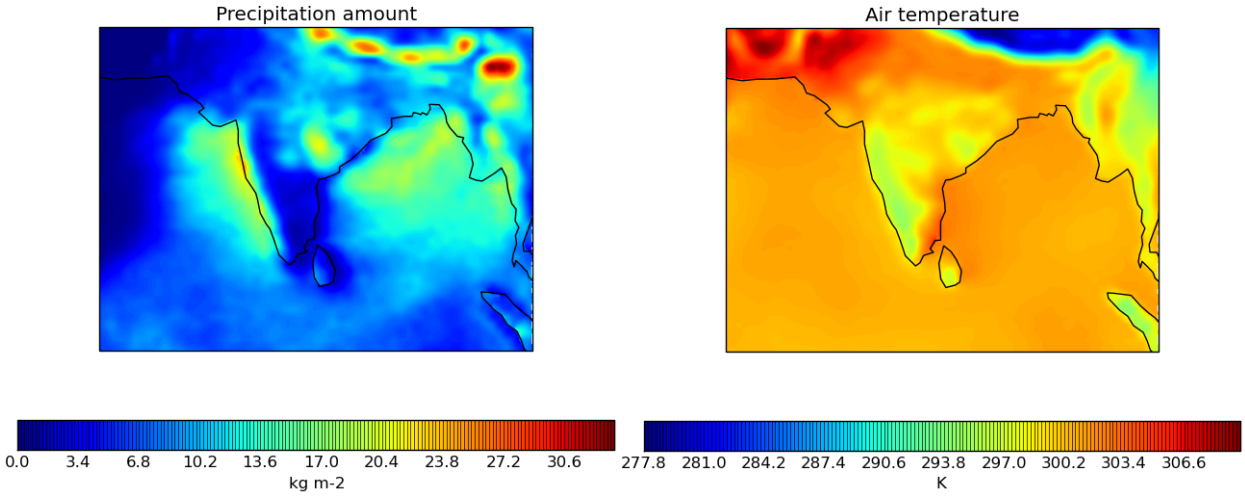

Figure 1: Mean daily precipitation (LHS) and 2m temperature (RHS) over India for July 2010 using ERA-Interim data.

Observation	Variables	Dates
Land SYNOP	Surface pressure, temperature, humidity, snow depth, wind	1978-2017
SYNOP cloud	cloud	1978-2017
Ship	Surface pressure, wind, temperature, humidity	1978-2017
Buoy	Surface pressure, wind, temperature	1979-2017
Sondes	Upper-air wind, temperature, humidity	1978-2017
PILOT and wind profilers	Upper-air wind	1978-2017
Aircraft	Flight-level temperature, wind	1978-2017
AIRS	AIRS radiances	2003-2017
ATOVS	HIRS/AMSU radiances	1998-2017
GPSRO	GPSRO bending angle	2006-2017
Ground GPS	Ground GPS slant delay	1999-2017
IASI	IASI radiances	2007-2017
satwinds/AMVs	wind	1988-2017
scatwinds	wind	1992-2017
SEVIRIclear	clear sky radiances	1982-2017
SSM/I	Total column water vapour, ocean wind speed	1987-2010
SSMIS	SSMIS radiances	2009-2017
TOVS	HIRS/MSU radiances	1979-2002

Table 1: List of observations expected to be assimilated into IMDAA. These include SYNOP - surface synoptic observations, AIRS - Atmospheric Infrared Sounder, (A)TOVS - (Advanced) Television Infrared Observation Satellites Operational Vertical Sounder, GPS - Global Positioning System, GPSRO - GPS Radio Occultation, IASI - Infrared Atmospheric Sounding Interferometer, SEVIRI - Spinning Enhanced Visible and Infrared Imager and SSMI(S) - Special Sensor Microwave Imager (/Sounder). AMVs are Atmospheric Motion Vectors, HIRS are High-resolution Infrared Radiation Sounders and (A)MSU are (Advanced) Microwave Sounding Units.

© Crown Copyright 2014

The system will cycle every six hours so that six hourly analysis fields are produced with hourly forecasts in between. The domain has yet to be agreed, but its horizontal resolution is likely to be of the order of $12km$ (or $0.11°$). The reanalysis will be produced on 70 model levels reaching to a height of approximately $80km$. Since the reanalysis is based on a full numerical weather prediction (NWP) system, a full set of physically consistent meteorological fields will be produced at each analysis and forecast time.

© Crown Copyright 2014

© Crown Copyright 2014

Name	Dates	Domain	Frequency	Horizontal Scale	Inputs & Method	Outputs
APHRODITE	1951-2007	60E-150E, 15S-55N	Daily	$30km/0.25°$	Estimated from surface station data	Precipitation
CMAP	1979-2011	Global	Monthly	$300km/2.5°$	Rain Gauge estimates merged with those of IR and PMW	Precipitation
CMORPH	2002-present	60N-60S	Sub-Daily	$8km/0.072°$	AMV propagated PMW precipitation estimates	Precipitation
CRU	1901-2012	Global	Monthly	$60km/0.5°$	Estimated from surface station data	Cloud, potential evaporation, precipitation, temperature, vapour pressure
GPCP	1979-present	Global	Monthly	$300km/2.5°$	Rain Gauge estimates merged with those of IR and PMW	Precipitation
GSMaP-MVK	2000-2010	Global	Sub-Daily	$11km/0.1°$	AMV propagated Kalman filtered PMW precipitation rates, adjusted by IR brightness temperature	Precipitation
NMSG	1998-2011	40E-118E, 39S-39N	Daily	$110km/1°$	Rain Gauge estimates merged with those of IR and PMW	Precipitation
TRMM	1997-present	50S-50N	Sub-Daily	$30km/0.25°$	Merged IR/PMW precipitation estimates, rescaled using rain gauges	Precipitation

Table 2: Gridded observation datasets: APHRODITE (Yatagai et al., 2012) - Asian Precipitation: Highly-Resolved Observational Data Integration Towards Evaluation of Water Resources, CMAP (Xie et al., 2007) - Climate Prediction Center's Analysis of Precipitation, CMORPH (Joyce et al., 2004) - Climate Prediction Center's Morphing Technique, CRU (Harris et al., 2013) - Climate Research Unit (University of East Anglia), GPCP (Huffman et al., 1997) - Global Precipitation Climatology Center, GSMap-MVK (Ushio and et. al, 2009) - Global Satellite Mapping of Precipitation moving vector with Kalman filter, NMSG (Mitra et al., 2003) - National Centre for Medium Range Weather Forecasting Merged Satellite and Gauges and TRMM (Huffman and et. al., 2007) - Tropical Rainfall Measuring Mission. AMVs are Atmospheric Motion Vectors derived from Infared (IR) satellite measurements and PMW are passive microwave measurements.

Met Office

Lateral boundary conditions will be taken from the global reanalysis ERA-Interim (Dee et al., 2011). The regional reanalysis aims to improve on the quality of representation of meteorological variables over this global dataset. Mean temperature and precipitation for July 2010 from ERA-Interim are shown in Figure 1. It is expected that the equivalent IMDAA data will be more detailed and more accurate.

3 Validation data

To validate a dataset a measure of truth, usually observations, is required. This proves problematic for reanalysis since most observations available at analysis time are assimilated into the system. The lack of suitable truth is particularly difficult in the tropics due to observation sparcity. However some independent observations are available and listed in Section 3.1. Reanalysis datasets can also be compared with other reanalyses, which are summarised in Section 3.2.

3.1 Independent observations

Table 2 lists gridded observation datasets against which the IMDAA reanalysis could be assessed. The datasets are based on surface station and satellite data. Although some surface station data will be assimilated into the reanalysis, the density of stations used for these gridded observation datasets is sufficiently dense that the reanalysis and observation data are reasonably independent. Precipitation data from satellites will not be assimilated.

Of the datasets listed in Table 2, APHRODITE is the best candidate for use as a truth since it provides daily data for a long period and at high resolution. The CRU dataset will also be valuable since it covers variables other than precipitation.

These gridded datasets are valuable as measures of truth against which to compare the reanalysis, however they should not be used to assess the data in isolation, but as part of a wide range of assessment activities since they do not cover a wide range of variables.

3.2 Reanalyses

There are several global reanalyses which cover the same time period and could be used for comparison with IMDAA over India. Table 3 summarises these datasets. The IMDAA reanalysis is expected to be of higher quality than these datasets for most variables since it will be produced at higher resolution than is affordable for global models. Therefore global reanalyses cannot be used as a measure of truth against which to compare the IMDAA reanalysis. They can, however, be used as a reference dataset to show the added value that the regional model produces. Comparison between the IMDAA reanalysis and global reanalyses will require an independent measure of truth.

Since the IMDAA reanalysis will be nested in ERA-Interim, this will be a natural choice as a reference dataset. ERA-Interim is reasonably high resolution, has had a wide range of uses (Iizumi et al. (2014); McCormick et al. (2014), etc.) and is considered a ‘gold-standard’ in global reanalyses (Wang et al. (2014), Hofer et al. (2012), etc.).

© Crown Copyright 2014

Finally, the South Asian Regional Reanalysis (SARR) project attempted to run a regional reanalysis over India using the Weather Research and Forecasting (WRF) model. There may be some overlapping periods of data that could be used for validation with IMDAA.

Reanalysis	**Centre**	**Horizontal scale**	**Model levels**	**Time span**
ERA-Interim (Dee et al., 2011)	ECMWF	$80km/0.7°$	60	1979-present
NCEP-R1 (Kalnay, 1996)	NCEP-NCAR	$300km/2.5°$	28	1948-present
JRA-55 (Ebita et al., 2011)	JMA	$60km/0.5°$	60	1958-present
MERRA (Rienecker and co authors, 2011)	NASA	$60km/0.5°$	72	1979-present

Table 3: Global reanalyses: ERA-Interim - European Centre for Medium Range Weather Forecasting (ECMWF) ReAnalysis Interim, NCEP-R1 - National Centers for Environmental Prediction (NCEP) Reanalysis 1, JRA-55 - Japanese 55-year Reanalysis and MERRA - Modern Era Retrospective Analysis For Research And Applications. NCAR is the National Center for Atmospheric Research, JMA is the Japanese Meteorological Administration and NASA is the National Aeronautics and Space Administration.

4 Assessment methods

Several key problems are known with numerical models when modelling the atmosphere over India. These will dominate the assessment of the reanalysis since they define how well the monsoon is represented in it. They are

1. Precipitation over India
2. Monsoon onset, peak, withdrawal and duration
3. Monsoon variability and teleconnection links

Methods for evaluating these aspects of the monsoon are discussed in Sections 4.1, 4.2 and 4.3, respectively. Reanalysis aims to provide physically consistent data across a wide-range of time and spatial scales for many atmospheric variables. Therefore methods for a more general assessment of the reanalysis are given in Section 4.4.

4.1 Precipitation

Precipitation is, perhaps, the most challenging variable to represent in an NWP system. In particular it is well known that it is not accurately represented by NWP over India (Bhowmik et al., 2007; Chunlei et al., 2014; Mannan et al., 2013). Validation of precipitation will therefore be a priority when assessing the quality of the IMDAA reanalysis. It is important that the model precipitation has accurate timing, location and intensity, and therefore these three aspects will be assessed. A temporal evaluation could involve comparing time series of reanalysis precipitation at specific locations (or across regions) with observations. The time-steps could be hourly, to assess the quality of the diurnal cycle, or daily/monthly, to assess seasonal variation. An example time-series is given in Figure 2. The time-series may be noisy and so clearer information may be obtained by

© Crown Copyright 2014

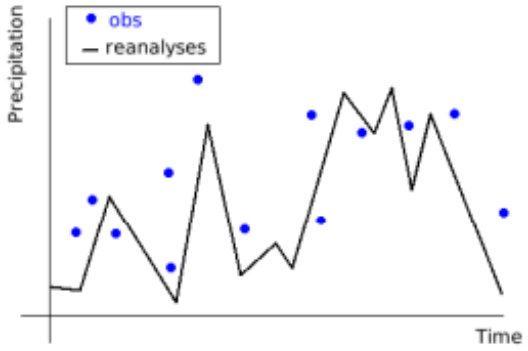

Figure 2: Example timeseries plot

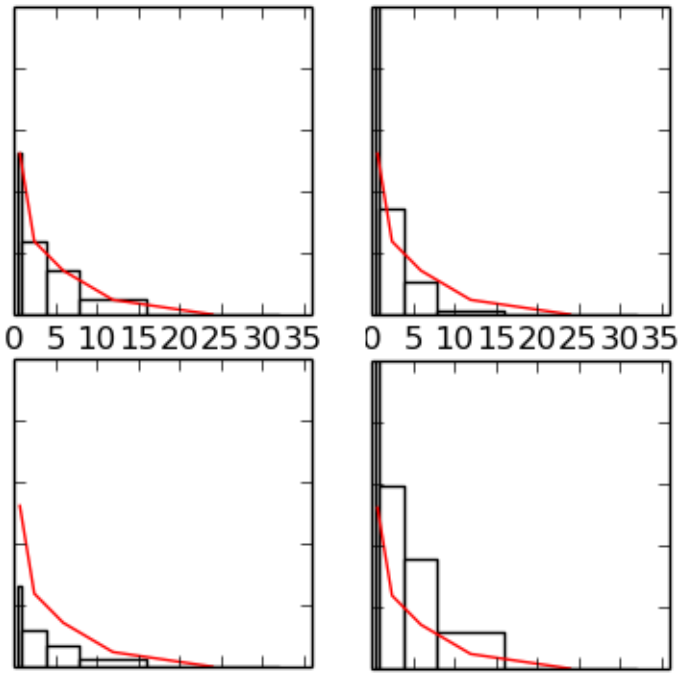

Figure 3: Example histograms. Lines are observed values, boxes are modelled values. Top row: LHS - model matches observations, RHS - model over-represents low threshold events and under-represents high threshold events. Bottom row: LHS - model drier than observed, RHS - model wetter than observed.

© Crown Copyright 2014

smoothing or averaging over a number of days/years. Ranade (2014) simplified further by categorising each day either as ‘wet’ or ‘dry’.

Quantitive representation of variables can be examined by comparing the precipitation distribution from the model with that of observations. (This could be extended to include variables such as number of dry or wet days, temperature, Tmin, Tmax etc.) Figure 3 shows example histograms in which the reanalysis distributions are compared to observation data. Mismatches of intensity between the two datasets can be easily identified. This diagnostic would be useful for comparison over a large-scale e.g. averaged over all of India or averaged over certain regions over the Indian land mass and over regions over the Indian Ocean.

The spatial distribution of mean seasonal climatologies of various measures can also be viewed in the form of spatial maps (Frei et al., 2003). Bias, relative error and/or correlation with observation maps will demonstrate how quality of representation of precipitation varies spatially. These maps may also be compared with those of global reanalyses to demonstrate where the regional reanalysis improves on these coarser models.

4.2 Monsoon onset, peak, withdrawal and duration

Knowledge of the date of onset of the monsoon is important for farmers across India (Webster et al., 1998). The capability of the IMDAA reanalysis to represent all aspects of past monsoons correctly is vital for improving representation of the monsoon in future forecast and dynamics studies, as well as producing an accurate statistical record over a long period. Vellinga et al. (2012) recently employed the use of three large-scale onset indicators to identify the monsoon onset over West Africa using MO seasonal hindcasts. The use of more than one criteria, in this way, for describing the monsoon onset gives robust results and improved understanding of the model performance. Broadly these indicators are

1. A large change in rainfall
2. A large change in outgoing longwave radiation
3. A circulation pattern indicative of the monsoon being established.

Using this criteria the onset dates across the region may be mapped and compared with observations and global reanalyses, as shown in Figure 4, demonstrating the ability of the model to represent this statistic. Similarly, the withdrawal, peak and duration of the monsoon could also be assessed using such methods and even pattern correlations computed between observations and the reanalysis for monsoon onset and withdrawal dates as in Sperber et al. (2013). Precipitation may also be plotted against date and against latitude, as in Figure 5. This allows a greater understanding of differences in the monsoon between datasets.

Figure 6 shows distibutions from a model of the three onset indicators and compares these against observation data. Assessment of distributions, as in Figure 6, will identify bias in the reanalysis.

4.3 Monsoon variability and teleconnection links

The South Asian monsoon tends to be dominated by two timescales of variability that have strong impacts on rainfall: (1) interannual variations sometimes associated with El Niño-Southern Oscillation (ENSO) and (2) 30-60 day intraseasonal oscillations (ISO) that modulate convection and embedded synoptic variations (e.g.

© Crown Copyright 2014

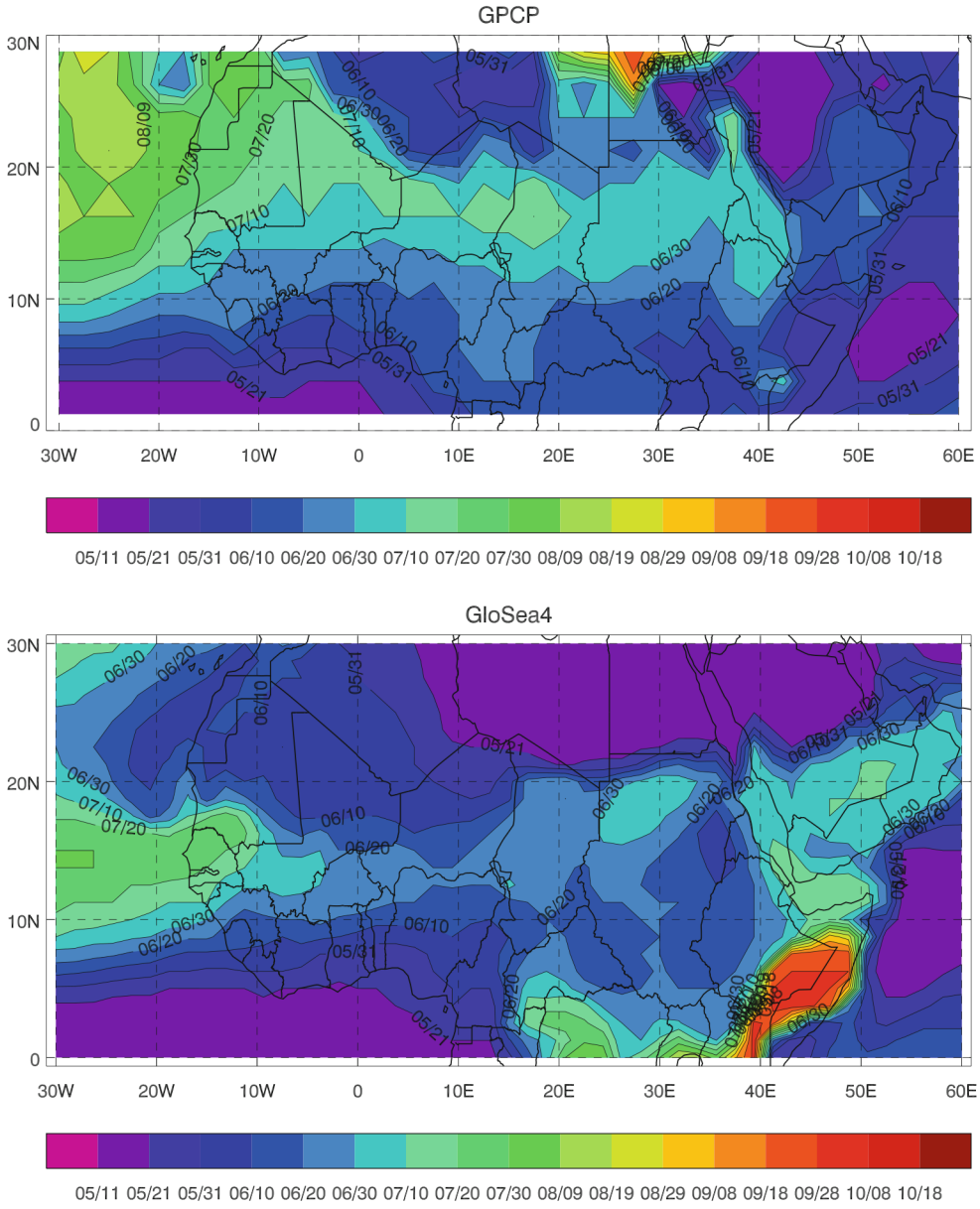

Figure 4: Example of isochrone plots showing climatology of monsoon onset dates for observations (top) and model (bottom). This is Figure 4 from Vellinga et al. (2012).

© Crown Copyright 2014

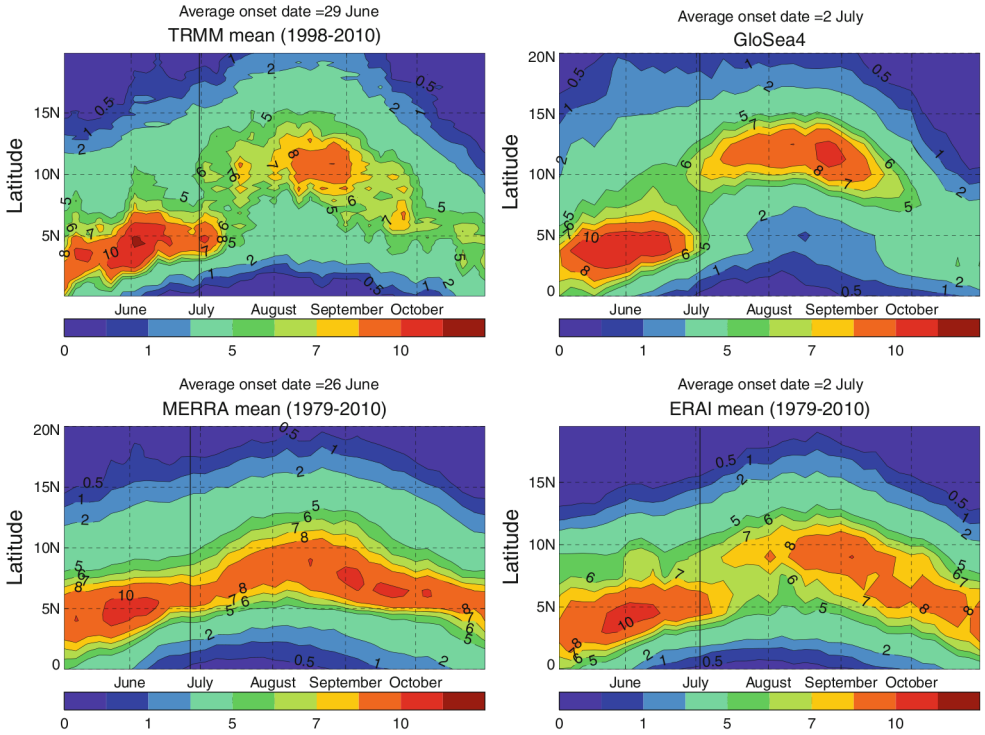

Figure 5: Example of time versus latitude plots of average pentad rainfall (mm/day), zonally averaged between 10°E-10°W for four datasets over the period 1998-2010. The mean monsoon onset date is indicated with a vertical line. This is Figure 1 from Vellinga et al. (2012).

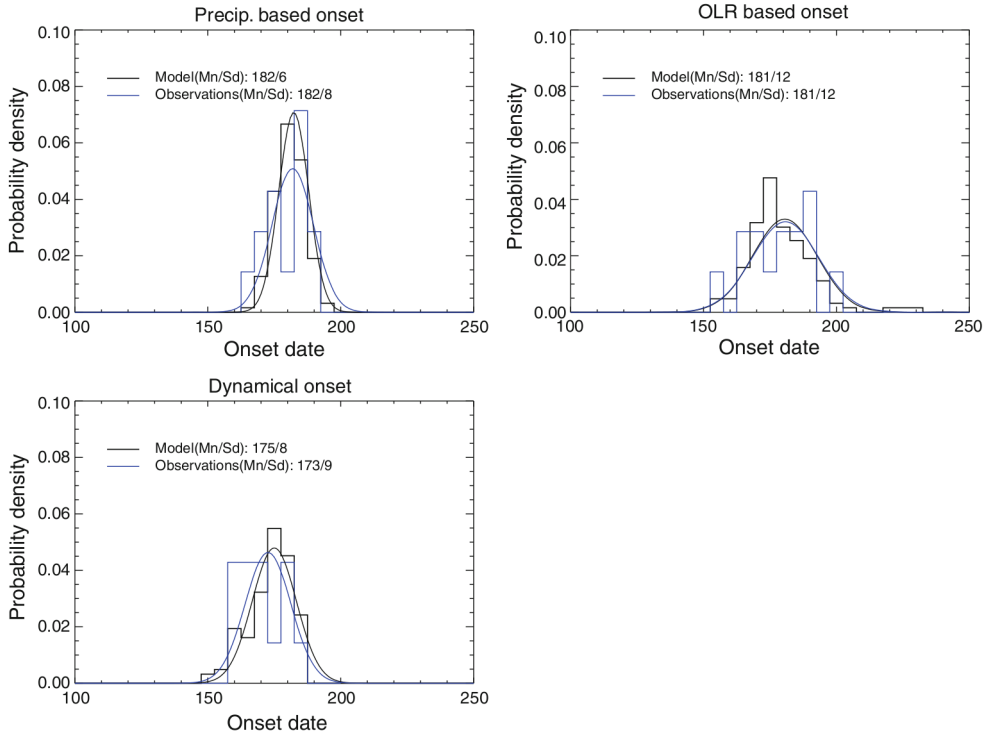

Figure 6: Example of frequency distributions of modelled (black) and observed (blue) monsoon onset dates using three different indicators. This is Figure 3 from Vellinga et al. (2012).

© Crown Copyright 2014

cyclones). ENSO is important as it can provide strong boundary forcings that establish remote teleconnections over the monsoon region (Achuthavarier et al., 2012; Ashrit et al., 2003) and the ISO can influence the onset and active-break cycles of the monsoon (Turner et al., 2011; Moon et al., 2012).

It is imperative that these phenomenon are accurately represented in the reanalyses for future studies. The strength of a teleconnection may be represented via correlations or composites. Correlations of the reanalysis data with established modes of variability known to affect the South Asian monsoon may be plotted and compared with similar data from observations and global reanalyses. Similarly average precipitation under each mode may be computed, i.e. composites. The difference between two composites also shows the strength of the teleconnection and, again, this can be compared with equivalent calculations for observation and global reanalysis data. Examples of correlation maps and composites are shown in Figures 7 and 8, respectively.

This validation technique will be useful in demonstrating the extent to which the regional model accurately captures global influences and may demonstrate that the model is useful for greater understanding of these teleconnections.

Additionally, the ability of the reanalysis to represent the variability of the monsoon should also be assessed. The year-to-year variation of the monsoon is usually monitored with objective measures of a representative variable (or variables) i.e. a dynamical monsoon index. There are are several such monsoon indices including; a large-scale measure of the development of the monsoon flow which takes into account both the development of the lower-level westerly jet and upper level easterlies (Webster and Yang, 1992), a relatively small scale index measuring reversal in the monsoon overturning Hadley circulation over the Indian region (Goswami et al., 1999), and a medium-scale measure of the low level re-circulation and development of the monsoon trough (Wang and Fan, 1999).

Subseasonal variability of the monsoon consists of dominant modes: an eastward and northward propagating boreal summer intraseasonal oscillation (BSIO) with a period of about 24-70 days and a westward propagating mode with a period of about 12-24 days (Lin et al., 2008). These subseasonal modes affect the onset and break of the monsoon and the formation, intensity and track of the tropical cyclones. Lin et al. (2008) use measures of filtered rainfall over the two different time periods to examine the subseasonal variability. The study examines a combination of measures which could be used for assessment of the reanalysis including looking at total sub-seasonal variance over the two modes of variability and lagged correlations to study the propagation of the modes. Figure 9 shows an example of a lag correlation of the 24-70 day precipitation anomaly for a set of observations and a model. This example demonstrates the model's inability to propagate past the Maritime Continent although the phase speed is realistic in this case. This type of analysis could be repeated for the reanalysis and compared against observations.

4.4 Reanalysis Quality

As well as demonstrating the ability of the reanalysis to accurately represent the monsoon, it is desireable for the reanalysis to accurately represent all meteorological variables on a sub-daily scale. Therefore it is useful to validate the reanalysis using forecast validation techniques. Such validation techniques can be either points-based, see Section 4.4.1, or spatial, see Section 4.4.2.

© Crown Copyright 2014

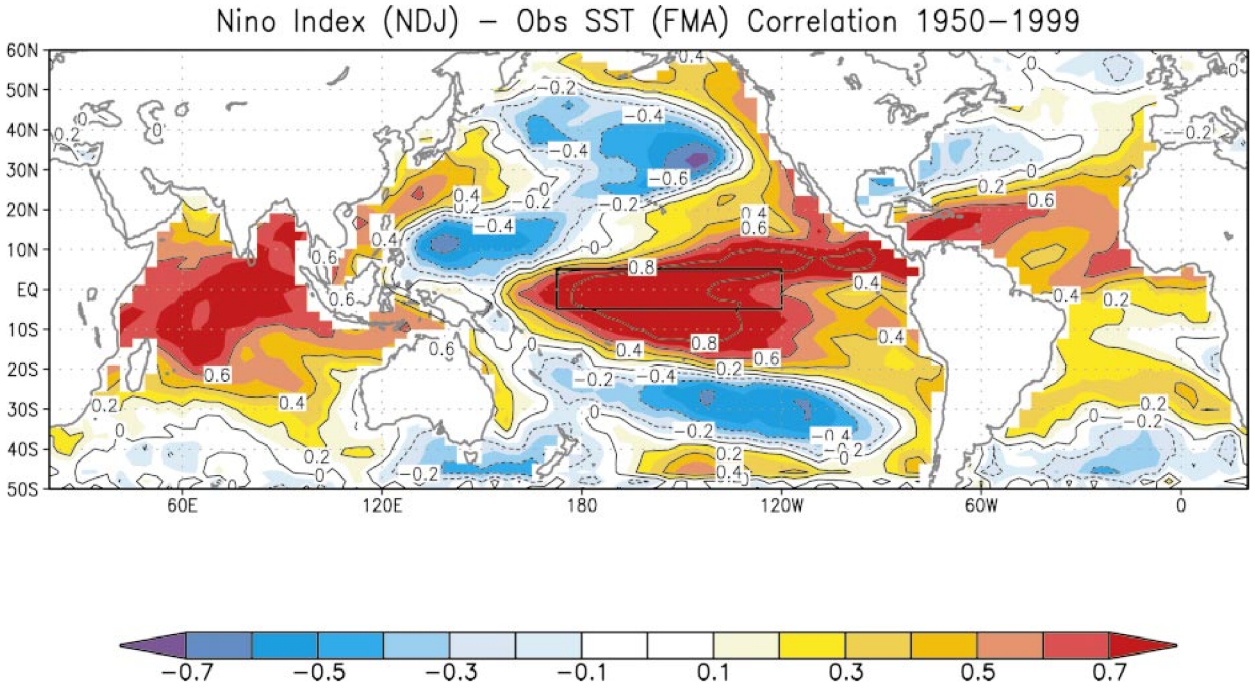

Figure 7: Example of correlation map showing correlation between sea surface temperature (SST) anomalies in the ENSO region (boxed) in November - January with global SST anomalies in the subsequent February - April. This is Figure 2 of Alexander et al. (2002).

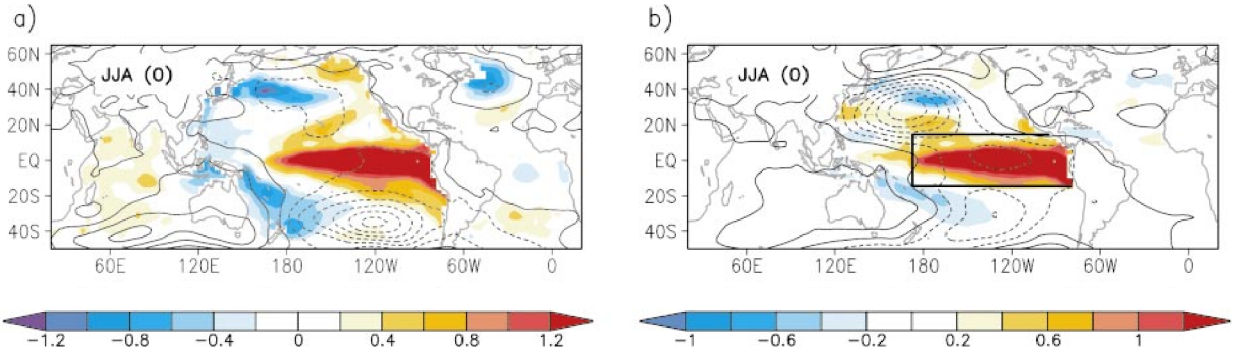

Figure 8: Example of composite showing (a) observed and (b) simulated El Nino - La Nina composite of surface pressure (contours) and sea surface temperature (shaded). This is part of Figure 6 of Alexander et al. (2002).

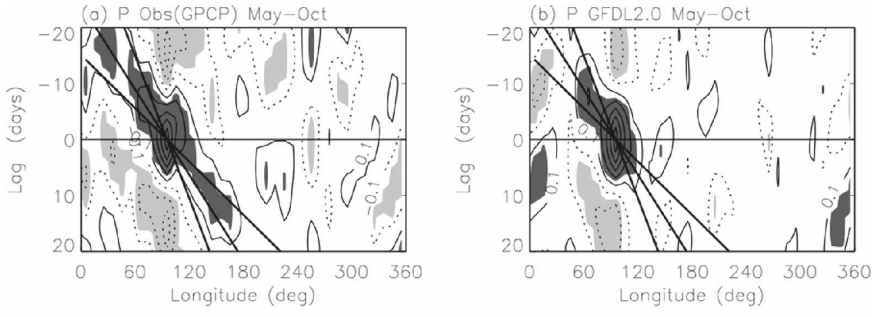

Figure 9: Example of a lag correlation of the 24-70 day precipitation anomaly averaged between 5 degrees and 25 degrees N with respect to itself at 15 degrees N, 95 degrees E: (a) observed and (b) an IPCC AR4 model. This is part of Figure 11 of Lin et al. (2008).

© Crown Copyright 2014

Field	Type	Weight
Temperature Surface	Skill	18
Temperature 850hPa	Skill	6
Temperature 500hPa	Skill	3
Wind Surface	Skill	18
Wind 500hPa	Skill	3
Relative Humidity 1000hPa	Skill	2
Relative Humidity 850hPa	Skill	1
Relative Humidity 500hPa	Skill	1
Cloud Cover 3/8	ETS	4
Cloud Cover 5/8	ETS	4
Cloud Cover 7/8	ETS	4
Precipitation 0.5mm	ETS	12
Precipitation 1mm	ETS	12
Precipitation 4mm	ETS	12
Total	-	100

Table 4: Regional reanalysis index of ETS and skill scores. Precipitation and Temperature are weighted most highly. Weights decrease with increasing height.

4.4.1 Points-based Validation

Since a good short range forecast implies a good analysis, the quality of the reanalysis may be assessed by measuring the quality of short forecasts. Following this method, observations to be assimilated may be used as truth since the observations at forecast time are independent of the previous reanalysis cycle. This allows a wide range of meteorological variables to be assessed. Since errors are expected to grow with forecast time, the scores do not give an absolute measure of quality, but at least give an upper limit on the errors in the reanalysis. To give a measure of the error of the reanalysis the quality of the short forecasts can be compared with a reference forecast, i.e. a global reanalysis.

At the Met Office NWP improvements are measured via global and UK indices, which are weighted sums of skill scores and equitable threat scores (ETS) for a number of useful meteorological variables. Such an index could be composed for the IMDAA reanalysis, mixing both smaller and larger scale variables. A suggested index is given in Table 4 which gives highest weight to wind, temperature and precipitation and whose weights increase with proximity to the surface. Skill scores and ETS are defined in Appendix A. Notable exceptions are pressure and geopotential height. These very large-scale variables are expected to be better represented in the global system and therefore are not included in this suggested index.

Bias can also be assessed by evaluating mean error for temperature, wind and relative humidity and frequency bias for precipitation and cloud cover.

Points-based validation allows assessment of the model at the highest resolution, but is not a like-for-like comparison since grid-box averages from the model are compared with point observations. This is particularly a problem for variables with large spatial variance such as precipitation. Grid-box averages are known to fail to accurately represent extreme events at short scales (Chen and Knutson (2008)).

Similar scores can be used to directly assess the reanalysis against independent datasets, e.g. APHRODITE. Using these scores with gridded observations is closer to a like-for-like comparison, but the scale of the assessment is limited by the resolution of the observation grid. Such assessments are very limited by which variables can be assessed, see Table 2.

© Crown Copyright 2014

4.4.2 Spatial Validation

Spatial validation techniques aim to address the 'double error' problem, where models are penalised twice for spatial error. This is particularly a problem when assessing precipitation since it has high spatial variance. These spatial validation techniques assess how well a model performs over an area covering multiple grid-boxes rather than at precise points. Instead of blindly penalising spatial displacements, such techniques register models as accurate above a certain scale. Therefore, at shorter scales, a model may score poorly, indicating that there are large errors at detailed scale and yet score well at larger scales, indicating that the dataset contains useful information.

A common tool for spatial validation is the Fractions Skill Score (FSS). The FSS compares the fraction of events across a domain in the model with that in the observations and is introduced in Roberts, N.M. and Lean, H.W. (2008) and described in Appendix B. Figure 10 is an example showing FSS of the two EURO4M reanalyses as relative differences to that of ERA-Interim for increasing scale over Germany.

FSS leads to the notion of 'useful' scales. These are spatial scales above which a model is better than a random model with the same proportion of events. It is desirable that IMDAA improves representation of precipitation during the monsoon both at larger scales and finer scales. This tool will be useful in demonstrating how the quality of representation varies with scale.

Spatial techniques can only assess data at scales at least several times the grid-box size of available gridded data. Therefore they are not very useful for providing information of the performance at very high resolution and are not very useful at assessing the quality of representation of the most extreme, small-scale events. Therefore to rigorously assess a precipitation dataset both spatial and points-based validation should be used.

Taylor diagrams allow evaluation of model data performance regarding the matching of spatial patterns using spatial correlation and standard deviation (Taylor, 2001). Spatial correlation demonstrates the degree to which the patterns match an observation dataset, while the standard deviation compares the amplitude of the variations. Taylor diagrams have been used to assess global precipitation in reanalyses (Bosilovich et al., 2008; Phillips and Gleckler, 2006) before by examining the spatial correlations and standard deviations from montly mean precipitation for a number of reanalyses. An example plot is shown in Figure 11 to demonstrate their ease of use for the evaluation of a variable. If the standard deviations are normalized to the observation dataset, then proximity to the coordinate $(1, 1)$ is a measure of the ability of the model to reproduce the spatial pattern. Use of Taylor diagrams can be used to assess precipitation patterns in IMDAA at a range of timescales.

5 Conclusions

The IMDAA regional reanalysis will be assessed both for how well it represents the monsoon and, more generally, for its quality across a wide range of atmospheric variables, time periods and spatial scales. Representation of the monsoon will focus on three key aspects that are known to be challenging to forecast models: precipitation, monsoon onset and teleconnections. The general assessment will employ validation techniques normally associated with NWP. Independent observations will be used as truth for these assessments. This is made possible either by using data which has largely not been assimilated into the reanalysis or by assessing short forecasts from the reanalysis. Global reanalyses make useful reference datasets to demonstrate the added value achieved from the regional system.

© Crown Copyright 2014

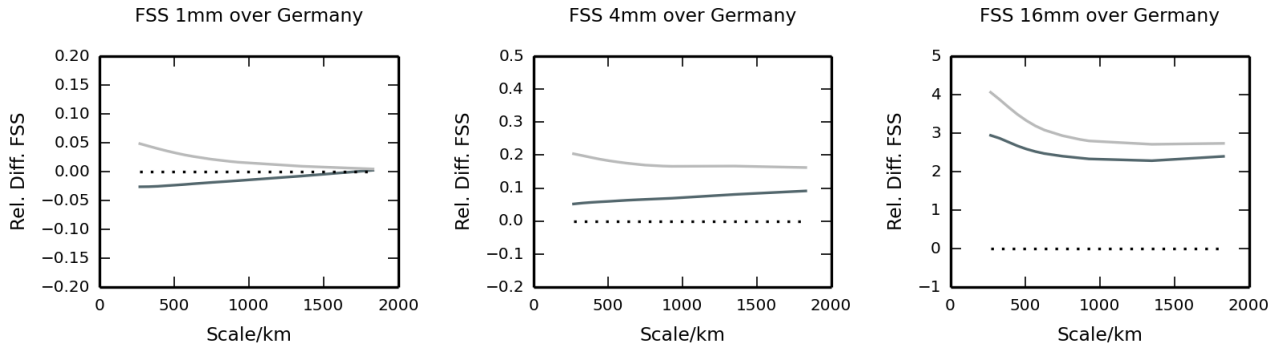

Figure 10: Monthly FSS averaged over 2008, difference relative to ERA-Interim, against increasing spatial scale for MO (light grey), HIRLAM (dark grey) and ERA-Interim (black, dotted). L-R thresholds 1mm, 4mm and 16mm over Germany. The accumulation period is twenty-four hours.

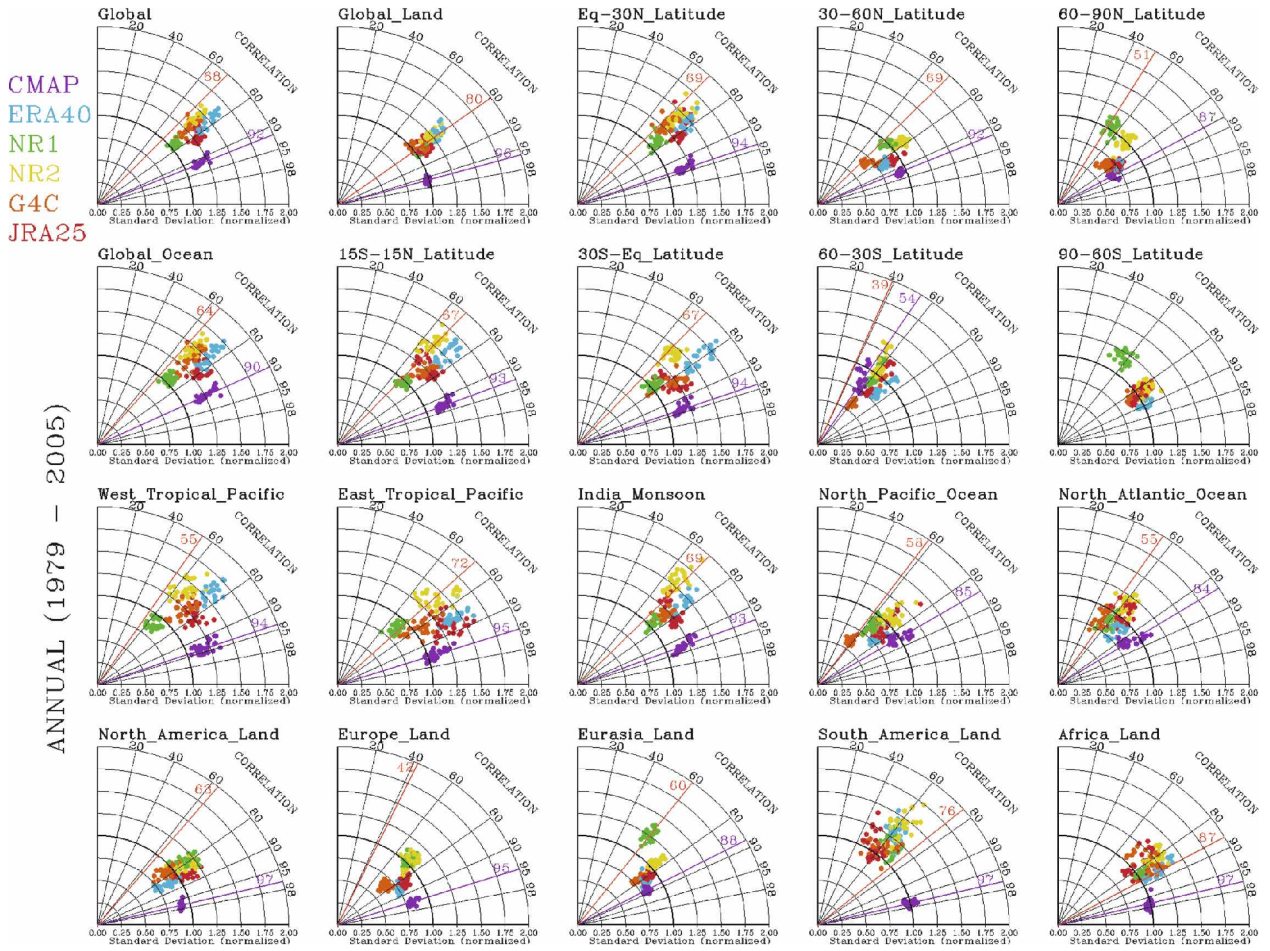

Figure 11: Example of Taylor diagrams for the annual mean correlations and standard deviations for various regions. The dots show correlations between different datasets and the observed values. This is Figure 3 from Bosilovich et al. (2008).

© Crown Copyright 2014

Further assessment methods may emerge once the data is available for investigation. The wide range of assessments proposed will demonstrate the ability of the reanalysis to represent both large and small scale features and will improve understanding of numerical modelling over the region.

© Crown Copyright 2014

A Points-Based Validation

For continuous data, such as temperature, it is natural to validate using RMSEs

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(m_i - o_i)^2} \tag{1}$$

where m and o are model and observation/truth values, respectively, for N casees. Bias can then be assessed via the mean error

$$\text{mean error} = \frac{1}{N}\sum_{i=1}^{N}(m_i - o_i)\,. \tag{2}$$

RMSEs do not take account of meteorological conditions, i.e. low errors are achieved during periods of high pressure, and so it is common to use a skill score which is the relative difference of the forecast RMSE and a persistence RMSE:

$$\text{skill} = \frac{RMSE^2_{\text{persistence}} - RMSE^2_{\text{forecast}}}{RMSE^2_{\text{persistence}}}. \tag{3}$$

This score cannot be used at analysis time, since the persistence will be identical to the forecast, but a similar score may be employed which uses say a global reanalysis as the reference error:

$$\text{skill} = \frac{RMSE^2_{\text{reference}} - RMSE^2_{\text{analysis}}}{RMSE^2_{\text{reference}}}. \tag{4}$$

A.1 Categorical

For less-continuous atmospheric variables such as precipitation, it is more appropriate to validate the data with a categorical score. The equitable threat score (ETS), or Gilbert score, is widely used to verify forecasts and was first proposed by Gilbert, G.K. (1884) to assess forecast events against observed events. The score arises from the following contingency table which counts the occurrences of each event in both the forecast and the observations as follows

-	observed	not observed
forecast	A	B
not forecast	C	D

The number of correct forecasts expected by chance is then the number of observed events multiplied by the number of forecast events divided by the total number of events, i.e.

$$A_r = \frac{(A+B)(A+C)}{A+B+C+D} \tag{5}$$

The number of correct forecasts minus those expected by chance is then

$$ETS^* = A - A_r \tag{6}$$

© Crown Copyright 2014

This is normalised to give a score of 1 when $B = 0$ and $C = 0$ (there are no incorrect forecasts) and to be penalised by large values of B and C.

$$ETS = \frac{A - A_r}{A + B + C - A_r} \tag{7}$$

$$ETS = \frac{AD - BC}{(B + C)n + AD - BC} \tag{8}$$

$$\text{where} \quad n = A + B + C + D \tag{9}$$

The associated frequency bias is simply the ratio of forecast events to observed events, i.e.

$$bias = \frac{A + B}{A + C}. \tag{10}$$

B Fractions Skill Score

The fractions skill score (FSS) was developed in Roberts, N.M. and Lean, H.W. (2008) to assess a model's skill in representing precipitation events at a given spatial scale. The FSS uses gridded observation data as truth and the scale assessed is based on $n \times n$ observation grid-boxes where n is a user-determined integer less than or equal to $2N - 1$ where N is the number of grid-boxes on the longest side of the observation grid.

The model data is reconfigured to observation resolution then, for a given threshold, the fraction of events modelled at scale $n \times n$ centred on grid-box i, j is given by

$$M_{ij}^n = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} I_M \left[i + k - 1 - \frac{n-1}{2}, j + l - 1 - \frac{n-1}{2} \right] \tag{11}$$

where $I_M[i, j]$ is an integer - 1 if there is an event at i, j and 0 otherwise. The observed fraction is similar, but a convolution kernel is applied to smooth the data to the appropriate scale. The kernel used here, following Roberts, N.M. and Lean, H.W. (2008), is a $n \times n$ mean filter, K_n, but any appropriate smoothing filter should produce similar results.

$$O_{ij}^n = \frac{1}{n^2} \sum_{k=1}^{n} \sum_{l=1}^{n} I_O \left[i + k - 1 - \frac{n-1}{2}, j + l - 1 - \frac{n-1}{2} \right] K_n(k, l) \tag{12}$$

The mean square error between observed and model fractions is then

$$MSE^n = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left(M_{ij}^n - O_{ij}^n \right)^2 \tag{13}$$

A reference error, which is the expected error from randomly rearranging the model's grid-boxes, is

© Crown Copyright 2014

$$MSE_r^n = \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left(M_{ij}^n\right)^2 + \frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} \left(O_{ij}^n\right)^2 \quad (14)$$

The fractions skill score is then given by

$$FSS^n = 1 - \frac{MSE^n}{MSE_r^n} \quad (15)$$

which is one minus the mean fraction error normalised by the largest possible fraction error obtainable using this scale, domain and fraction of events. Useful skill is achieved when the FSS is at least $0.5(1 + f_0)$ where f_0 is the frequency of observed events.

© Crown Copyright 2014

References

D. Achuthavarier, V. Krishnamurthy, B. P. Kirtman, and B. Huang. Role of the indian ocean in the ENSO-Indian summer monsoon teleconnection in the NCEP climate forecast system. *J. Climate*, 25:2490–2508, 2012.

M. A. Alexander, I. Blade, M. Newman, J. R. Lanzante, N.-C. Lau, and J. D. Scott. The atmospheric bridge: The influence of ENSO teleconnections on air-sea interaction over the global oceans. *Bull. Amer. Meteor. Soc.*, 15:2206–2231, 2002.

R. G. Ashrit, H. Douville, and K. R. Kumar. Response of the Indian monsoon and ENSO-monsoon teleconnection to enhanced greenhouse effect in the CNRM coupled model. *Journal of the Meteorological Society of Japan. Ser. II*, 81:779–803, 2003.

S. K. Roy Bhowmik, D. Joarder, and H. R. Hatwar. Evaluation of precipitation prediction skill of IMD operational NWP system over Indian Monsoon region. *Meteorology and Atmospheric Physics*, 95:205–221, 2007.

M. G. Bosilovich, J. Chen, F. R. Robertson, and R. F. Adler. Evaluation of Global Precipitation in Reanalyses. *J. Appl. Meteor. Climatol.*, 47:2279–2299, 2008.

C-T. Chen and T. Knutson. On the verification and comparison of extreme rainfall indices from climate models. *J. Climate*, 21:1605–1621, 2008. doi: 10.1175/2007JCLI1494.1.

L Chunlei, R. P. Allan, M. Brooks, and S. Milton. Comparing tropical precipitation simulated by the met office nwp and climate models with satellite observations. *J. Appl. Meteor. Climatol.*, 53:200–214, 2014.

T. Davies, M. J. P. Cullen, A. J. Malcolm, M. H. Mawson, A. Staniforth, A. A. White, and N Wood. A new dynamical core for the Met Office's global and regional modelling of the atmosphere. *Q. J. R. Meteorol. Soc.*, 131:1759–1782, 2005.

D. P. Dee, S. M. Uppala, A. J. Simmons, P. Berrisford, P. Poli, S. Kobayashi, U. Andrae, M. A. Balmaseda, G. Balsamo, P. Bauer, P. Bechtold, A. C. M. Beljaars, L. van de Berg, J. Bidlot, N. Bormann, C. Delsol, R. Dragani, M. Fuentes, A. J. Geer, L. Haimberger, S. B. Healy, H. Hersbach, E. V. Hlm, L. Isaksen, P. Kllberg, M. Khler, M. Matricardi, A. P. McNally, B. M. Monge-Sanz, J.-J. Morcrette, B.-K. Park, C. Peubey, P. de Rosnay, C. Tavolato, J.-N. Thpaut, and F. Vitart. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656):553–597, 2011. ISSN 1477-870X. doi: 10.1002/qj.828. URL `http://dx.doi.org/10.1002/qj.828`.

A. Dobler and B. Ahrens. Four climate change scenarios for the Indian summer monsoon by the regional climate model COSMO-CLM. *J. Geophys. Res.*, 116:1–13, 2011.

A. Ebita, S. Kobayashi, Y. Ota, M. Moriya, R. Kumabe, K. Onogi, Y. Harada, S. Yasui, K. Miyaoka, K. Takahashi, H. Kamahori, C. Kobayashi, H. Endo, M. Soma, Y. Oikawa, and T. Ishimizu. The Japanese 55-year reanalysis JRA-55: an interim report. *SOLA*, 7:149–152, 2011.

C. Frei, J. H. Christensen, M. deque, D. Jacob, R. G. Jones, and P. L. Vidale. Daily precipitation statistics in regional climate models: Evaluation and intercomparison for the European Alps. *J. Geophys. Res.*, 108:4124, 2003.

© Crown Copyright 2014

S. Gadgil. The indian monsoon and its variability. *Ann. Rev. Earth Planet Sci.*, 31:429–467, 2003.

Gilbert, G.K. Finley's tornado predictions. *Amer. Meteor. J.*, 1:166–172, 1884.

B. N. Goswami. South Asian Summer Monsoon: An Overview. *Report of the International Committee of the Third International Workshop on Monsoons*, 70, 2004.

B. N. Goswami, V. Krishnamurti, and H. Annamalai. A broad scale circulation index for the interannual variability of the indian summer monsoon. *Q. J. R. Meteorol. Soc.*, 125:611–633, 1999.

I. Harris, P. D. Jones, T. J. Osborn, and D. H. Lister. Updated high-resolution grids of monthly climatic observations - the CRU TS3.10 dataset. *Int. J. Climatol.*, 34:623–642, 2013.

M. Hofer, B. Marzeion, and T. Mlg. Comparing the skill of different reanalyses and their ensembles as predictors for daily air temperature on a glaciated mountain (Peru). *Climate Dynamics*, 39(7-8):1969–1980, 2012. doi: 10.1007/s00382-012-1501-2.

G. J. Huffman and et. al. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined sensor precipitation estimates at fine scales. *J. Hydrometeorol.*, 8:38–55, 2007.

G. J. Huffman, R. F. Adler, P. Arkin, A. Chang, R. Ferraro, A. Gruber, J. Janowiak, A. McNab, B. Rudolf, and U. Schneider. The global precipitation climatology project (GPCP) combined precipitation dataset. *Bull. Amer. Meteor. Soc.*, 78:5–20, 1997.

T. Iizumi, M. Okada, and M. Yokozawa. A meteorological forcing data set for global crop modeling: Development, evaluation, and intercomparison. *J. Geophys. Res. Atmos.*, 119:363–384, 2014. doi: 10.1002/2013JD020130.

R. Joyce, J. E. Janowiak, P. A. Arkin, and P. Xie. CMORPH: A method that produces global precipitation estimates from passive microwave and infrared data at high spatial and temporal resolution. *J. Hydrometeorol.*, 5:487–503, 2004.

E. Kalnay. The NCEP/NCAR 40-year reanalysis project. *Bull. Amer. Meteor. Soc.*, 77:–33, 1996.

J. Lin, K. M. Weikman, G. N. Kiladis, B. E. Mapes, S. D. Schubert, M. J. Suarez, J. T. Bacmeister, and M. Lee. Subseasonal variability associated with asian summer monsoon simulated by 14 IPCC AR4 coupled GCMs. *J. Clim*, 21:4541–4567, 2008.

M. A. Mannan, M. A. M. Chowdhury, and S. Karmakar. Application of {NWP} model in prediction of heavy rainfall in bangladesh. *Procedia Engineering*, 56:667–675, 2013.

B. T. McCormick, M. Herzog, J. Yang, M. Edmonds, T. A. Mather, S. A. Carn, S. Hidalgo, and B. Langmann. A comparison of satellite- and ground-based measurements of SO2 emissions from Tungurahua volcano, Ecuador. *J. Geophys. Res. Atmos.*, 119:4264–4285, 2014. doi: 10.1002/2013JD019771.

A. K. Mitra, M. Das Gupta, S. V. Singh, and T. N. Krishnamurti. Daily rainfall for the indian monsoon region from merged satellite and rain gauge values: Large-scale analysis from real-time data. *J. Hydrometeor*, 4: 769–781, 2003.

© Crown Copyright 2014

J.-Y. Moon, B. Wang, K.-J. Ha, and J.-Y. Lee. Teleconnections associated with Northern Hemisphere summer monsoon intraseasonal oscillation. *Clim Dyn*, 40:2761–2774, 2012.

T. J. Phillips and P. J. Gleckler. Intercomparison of four global precipitation data sets and their correlation with increased eurasian river discharge to the arctic ocean. *J. Geophys. Res.*, 111:D21112, 2006.

Ranade. A verification of spatio-temporal monsoon rainfall variability across Indian region using NWP model output. *Meteorol. Atmos. Phys*, 125:43–61, 2014.

F. Rawlins, S.P. Ballard, K.J. Bovis, A.M. Clayton, D. Li, G.W. Inverarity, A.C. Lorenc, and T.J. Payne. The met office global 4-dimensional data assimilation system. *Q. J. R. Met. Soc.*, 133:347–362, 2007.

M. M. Rienecker and co authors. Merra - nasa's modern-era retrospective analysis for research and applications. *J. Climate*, 24:3624–3648, 2011.

Roberts, N.M. and Lean, H.W. Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. *Mon. Wea. Rev.*, 136:78–97, 2008.

K. R. Sperber, H. Annamalai, I.-S. Kang, A. Kitoh, A. Moise, A. Turner, B. Wang, and T. Zhou. The Asian summer monsoon: an intercomparison of CMIP5 vs. CMIP3 simulations of the late 20th century. *Clim. Dyn.*, 41:2711–2744, 2013.

K. E. Taylor. Summarizing multiple aspects of model performance in a single diagram. *J. Geophys. Res.*, 106: 7183–7192, 2001.

A. Turner, K. R. Sperber, J. Slingo, G. Meehl, C. R. Mechoso, M. Kimoto, and A. Giannini. Modelling Monsoons: Understanding and predicting current and future behaviour. *Global Monsoon System: Research and Forecast*, 2011.

A. G. Turner and H. Annamalai. Climate change and the South Asian summer monsoon, model projections. *Nature Climate Change*, 2:587–595, 2012.

T. Ushio and et. al. A Kalman filter approach to the Global Satellite Mapping of Precipitation (GSMaP) from combined passive microwave and infrared radiometric data. *J. Meteorol. Soc. Jpn.*, 87:137–151, 2009.

M. Vellinga, A. Arribas, and R. Graham. Seasonal forecasts for regional onset of the West African monsoon. *Clim. Dyn.*, 40:3047–3070, 2012.

B. Wang and Z. Fan. Choice of South Asian Summer Monsoon Indices. *Bull. Am. Meteorol. Soc.*, 80:629–638, 1999.

S. Wang, M. Zhang, M. Sun, B. Wang, X. Huang, Q. Wang, and F. Feng. Comparison of surface air temperature derived from NCEP/DOE R2, ERA-Interim, and observations in the arid northwestern China: a consideration of altitude errors. *Theoretical and Applied Climatology*, pages 1–13, 2014. doi: 10.1007/s00704-014-1107-1.

P. J. Webster and S. Yang. Monsoon and ENSO: selectively interactive sytems. *Q. J. R. Meteorol. Soc.*, 118: 877–926, 1992.

© Crown Copyright 2014

P. J. Webster, V. O. Magana, T. N. Palmer, J. Shukla, R. A. Tomas, M. Yanai, and T. Yasunari. Monsoons: Processes, predictability, and the prospects for prediction. *J. Geophys Res.*, 103:14451–14510, 1998.

P. Xie, P. A. Arkin, and J. E. Janowiak. CMAP: The CPC merged analysis of precipitation. *Advances In Global Change Research*, 28:319–328, 2007.

A. Yatagai, K. Kamiguchi, O. Arakawa, A. Hamada, N. Yasutomi, and A. Kitoh. APHRODITE: Constructing a long-term daily gridded precipitation dataset for asia based on a dense network of rain gauges. *Bull. Amer. Meteor. Soc.*, 93:1401–1415, 2012.

© Crown Copyright 2014

Met Office
FitzRoy Road, Exeter
Devon, EX1 3PB
UK

Tel: 0870 900 0100
Fax: 0870 900 5050
enquiries@metoffice.gov.uk
www.metoffice.gov.uk